CHARLES UNIVERSITY, FACULTY OF SCIENCE
DEPARTMENT OF PHYSICAL AND MACROMOLECULAR CHEMISTRY
ALBERTOV 6, 128 43 PRAGUE 2
CZECH REPUBLIC

Prague, June 2nd, 2021

Opponent's review of the doctoral thesis by Mgr. Vjačeslav Treťjačenko entitled

„The effect of amino acid repertoire on protein structure evolution"

The doctoral thesis presented by Vjačeslav Treťjačenko focuses on the investigation of how most recent additions to the amino acid alphabet affected the protein structure/function relationship and on the properties of random proteins as the evolutionary point-zero for the earliest sequences as well as for proteins emerging de novo from the non-coding parts of the genome. Random proteins are of a special interest as they unveil the unexposed side of the protein sequence space. Therefore the aim of this doctoral thesis is important and timely.

The "Results and Discussion" section of the thesis is divided into four parts (aims). In the first one, Vjačeslav Treťjačenko generated 4 datasets with 10,000 random protein sequences of 100 amino acid residues long, performed prediction of their secondary structure, disorder and tendency to aggregate and compared them with the naturally occurring proteins. Subsequently, 45 sequences from the random dataset were expressed, purified and biophysically characterized. This study revealed that the random proteins exhibit similar secondary structure content and overall aggregation propensity as natural proteins, but the naturally occurring proteins show better optimization towards aggregation suppression while maintaining comparable structure content. Furthermore, this study also showed that random proteins exhibit low aggregation, high solubility, and reasonable intracellular tolerance to be suitable precursors for the following evolutionary optimization as novel proteins.

The second part of the Results section deals with the development of combinatorial library design tool (CoLiDe). The purpose of this tool is to compute such a combination of degenerate codons which, when combined into one DNA template, will produce a protein-coding library with user defined amino acid ratios.

The third part of the Results section focuses on the characterization of combinatorial protein libraries with distinct amino acid alphabets. Vjačeslav Treťjačenko used the previously developed CoLiDe tool to design two libraries with 20 (canonical set) and 10 (early set) amino acid alphabets, prepared these proteins and studied their behaviour in the presence of contemporary protein folding enhancers (DnaK, DnaJ and GrpE). In particular, he studied their solubility, aggregation propensity and sensitivity against two different proteases. This part revealed different structural tendencies within the random sequence space.

The last part of the Results section deals with the characterization of aromatic-less variant of dephospho coenzyme-A kinase (DPCK). This study confirmed the role of aromatic amino

acids in achieving the structural stability of contemporary proteins but it also demonstrated that the enzyme is active even in their absence.

The presented doctoral thesis is written in a shortened version and its formal structure follows the standard division. Both the formal and the graphical quality of the presented thesis are satisfactory. However, in the section "Results and Discussion" I lacked a greater discussion of the obtained results, especially in the context of similar studies from the literature. The quality of English, at least as I can judge, is acceptable. Obtained results were published in three papers in international journals with IF, Vjačeslav Treťjačenko is the first author on two of them.

For the sake of discussion during the defense of the thesis, I have following questions related to presented results:

1. The early amino acid alphabet consisted of G, A, D, E, V, I, L, P, T, and S, thus no positively charged residues were present. Is there any explanation for that?
2. Is it correct to assume that the earliest (very first) proteins composed of just G, or G,A, or G,A,D (based on suggested relative age)? Are there any studies on such proteins?
3. Is there any theory/hypothesis why we have only 21 (22) proteinogenic amino acids, which are genetically encoded? Why the evolution stopped at this number?
4. Why the presence of DnaK suppressed the expression of 10E library?
5. Why His residues were not mutated in DPCK-LH variant (compared to DPCK-M)? Can this be the reason for the better structural stability of DPCK-LH compared to the DPCK-M variant?

In conclusion, the doctoral thesis presented by Vjačeslav Treťjačenko represents a significant contribution to the characterization of properties of evolutionary early proteins, especially with respect to prebiotically plausible amino acids. The thesis is written in intelligible language and obtained results were published in international journals with IF.

In conclusion, the presented thesis clearly demonstrates that Vjačeslav Treťjačenko is able of independent scientific work. Since the presented thesis satisfies all requirements for the doctoral thesis **I fully recommend its acceptance**.

Prof. RNDr. Tomas Obsil, Ph.D.
Tel.: +420-221 951 303
E-mail: obsil@natur.cuni.cz